DOCUMENT RESUME

ED 463 330		TM 033 762
AUTHOR	Wang, Jianjun	
TITLE	An In-depth Analysis of Achievement Gaps between 7th and 8th Grades in the TIMSS Database.	
SPONS AGENCY	California State Univ., Bakersfield.	
PUB DATE	2002-04-00	
NOTE	18p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 1-5, 2002). Supported by the CSUB Teaching Facility and Learning Center and a TNT grant.	
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)	
EDRS PRICE	MF01/PC01 Plus Postage.	
DESCRIPTORS	*Academic Achievement; Foreign Countries; International Studies; *Junior High School Students; Junior High Schools; *Mathematics Achievement; *Scores; Test Construction; *Test Items	
IDENTIFIERS	Statistical Analysis System; *Third International Mathematics and Science Study	

ABSTRACT

Middle school data from the Third International Mathematics and Science Study (TIMSS) are analyzed in this study to compare achievement differences between adjacent grades. To facilitate this computer-based analysis, a Statistical Analysis System (SAS) program has been developed to transpose matrices of the item scores in more than 40 countries. The results indicate that not all TIMSS items have a higher mean score at the upper grade in each nation. Features of item construction are discussed to disentangle the issue of reflecting grade gaps in mathematics and science achievement. These analyses may help enrich understanding of other comparative studies using the TIMSS benchmark. (Contains 5 tables and 15 references.) (Author/SLD)

Running head: TIMSS Data Analysis

An In-depth Analysis of Achievement Gaps Between 7th and 8th Grades in the TIMSS Database*

Jianjun Wang

California State University, Bakersfield

*Paper presented at the 2002 annual meeting of the American Educational Research Association (AERA), New Orleans, April 1-5. Support for this project was supported in part by the CSUB Faculty Teaching and Learning Center and a TNT grant in collaboration with Mr. Doug Gardner. The opinions expressed in this manuscript are solely those of the author, and do not necessarily represent these partners and organizations.

An In-depth Analysis of Achievement Gaps

Between 7th and 8th Grades in the TIMSS Database

Abstract

Middle school data from the Third International Mathematics and Science Study (TIMSS) are analyzed in this study to compare achievement difference between adjacent grades. To facilitate this computer-based data analysis, a SAS program has been developed to transpose matrices of the item scores in more than forty countries. The results indicate that not all TIMSS items have a higher mean score at the upper grade in each nation. Features of the item construction have been discussed to disentangle the issue of reflecting grade gaps in mathematics and science achievement. These analyses may help enrich understanding of other comparative studies using the TIMSS benchmark.

An In-depth Analysis of Achievement Gaps

Between 7th and 8th Grades in the TIMSS Database

The Third International Mathematics and Science Study (TIMSS) is the largest and most ambitious project in comparative education. In the released data from middle schools, TIMSS researchers gathered student scores at 7th and 8th grades from around forty countries. One of the key factors behind the achievement difference is the increase of learning experience between these adjacent grades that accounts for 12.5% of students' school life. Given the average scores of a test item, students at the upper grade are generally expected to outperform their peers at the lower grade in each nation. To date, this pattern of score improvement has not been confirmed by TIMSS findings at the item level. The purpose of this investigation is to examine the item score difference between adjacent grades in each of the TIMSS participating nations. As more educators consider using the TIMSS benchmark to evaluate school effectiveness (Martin, Mullis, et al., 1998; Mullis, Martin, et al., 1998), results of this study may facilitate interpretation of the assessment results in an international context.

Literature Review

In the existing TIMSS reports, item scores have been aggregated into total scores to reflect the overall mathematics and science performance in each nation. Within each subject, subcategories were developed to group TIMSS items in specific content domains (Beaton, Martin, et al., 1996; Beaton, Mullis, et al., 1996). Still, some researchers recognized limitations of the aggregated total scores. Schmidt, McKnight, Cogan,

Jakwerth, and Houang (1999) observed,

> TIMSS achievement reporting thus far has been limited to global mathematics and science scale scores and to reporting the national percentages of items correct in a set of six 'reporting categories' in both subjects. These reporting categories were still so broad - as the global scores obviously were - as to include somewhat disparate items. (p. 117)

Beyond the designated categorizations for TIMSS reporting, more specific investigations can be conducted on test scores at the item level. Regardless of the contextual differences among various nations, it is incomprehensible to observe a drop of academic achievement on the same set of test items as students move from a lower grade to a higher grade within that country. Although TIMSS is not a longitudinal study, the average difference in academic performance can be employed to measure the cross-sectional gap between the adjacent grades in each nation. If the test content is covered by a curriculum at the upper grade, the results will show an increase in academic achievement. In addition, maturation and cognitive development are also in favor of the senior students, causing higher scores at the upper grade (Peterson, 1986; Walker & Madhere, 1987). Whereas school curricula may vary across different countries, the between-grade comparison is made within each country, and thus, the item performance can be linked to the domestic condition of science and mathematics education. Schmidt, et al. (1999) asserts that "it is precisely these content-specific differences among items that make achievement assessments curricularly sensitive" (p. 116). In this regard, the in-depth analysis of item scores may not only help interpret empirical measures of the grade

difference within each country, but also facilitate a comparison of the curricular emphases among different nations.

Methods

A straightforward approach to checking the score difference is to subtract item mean scores between the adjacent grades. If this issue involves only one item or a few items, the subtraction can be easily completed through hand-calculations. Unfortunately, the TIMSS instrument includes 429 multiple-choice, 43 short-response, and 29 extended-response items (Lange, 1997), requiring more than 501 comparisons for each nation. To cover the comparisons of item performance across multiple grades in around forty countries, the overall computing operation involves a total of 21,042 subtractions. This number does not include Israel and Kuwait due to their single-grade participation in the TIMSS middle-school investigation. Other countries like Sweden and Switzerland have gathered data from three adjacent grades, and thus, demand more effort on the computation. Without a computer program, no researchers have made the detailed comparisons of item scores in all countries (http://www.timss.org/timss1995i/Items. html).

A technical difficulty of the statistical computing hinges on the existing data structure. Using standard statistical software packages, such as SPSS or SAS, one can easily obtain the item mean scores for each grade. The results may be exported into a new database that contains the information of country, grade, and item performance. However, because the mean scores from different grades and countries are listed under variables in various columns (see Table 1), no statistical procedures in SAS or SPSS can

be employed to systematically calculate the mean score difference between adjacent rows of the same column (personal communication with technical consultants at SAS and SPSS, June 14, 2001).

After consulting with statisticians of several software companies and TIMSS experts at Boston College, three steps have been taken to re-organize the mean-score database for statistical computing. In the first step, the following SAS codes are employed to transpose the list of variable names in Table 1 into a column in Table 2:

```
proc transpose data=TIMSS95 out=new;
by n idcntry idgrader;
var BSMMA01 BSMMA02 BSMMA03 … ;
```

Insert Tables 1 & 2 around here

In the second step, a LAG function is introduced to calculate the mean score difference. This function is available in both SPSS and SAS. SPSS (1988) clarified, "PREV4=LAG (GNP, 4) returns the value of GNP for the fourth case before the current one" (p. 122). The equivalent statement in SAS is PREV4=LAG4(GNP) (SAS Institute, 1990). Accordingly, assuming <u>col1</u> to be the variable of item mean scores for each grade in each nation, LAG(*col1*) contains data that have one-case lag from data in the original <u>col1</u> column. The score subtraction between adjacent grades can be completed by an internal DIF function equivalent to [col1-LAG(col1)] (SAS Institute, 1990), i.e., mean_diff = dif(col1) (Table 3).

Inevitably, because mean scores are arranged by *grade* and *country*, the *mean_diff* computing also includes score subtractions between the last record of the previous country and the first record of the next country. In the transposed data structure (Table 2), these records are linked to different items in adjacent countries. As scores from different items often deal with different tasks, a lower grader in Japan may not necessarily score lower than a higher grader from South Africa on different test items. In the third step, a SAS command "if first.idcntry then mean_diff=.;" is issued to remove the subtraction results across the country borders. In addition, a statement "if mean_diff < 0;" is employed to select these items that have resulted in higher average scores at the lower grade in each country. The entire SAS program is gathered in Table 3 to implement the three-step approach. Because the LAG function is also available in SPSS, this approach can be readily adapted in SPSS-based TIMSS data analyses.

Insert Table 3 around here

Results

Before starting the data analysis, the author first calculated item scores to ensure that the results match the released item scores in TIMSS reports (Beaton, Martin, et al., 1996; Beaton, Mullis, et al., 1996). The transpose of the data matrix was completed through collaboration with a SAS consultant (SAS Tracking Number: us5512465). The SASLOG file has been examined to confirm that there were no syntax mistakes in the computer program. A portion of the final output has been copied in Table 4 for

illustration. Results of the comprehensive item analysis for all participating countries are assembled in Table 5.

Insert Tables 4 & 5 around here

Inspection of Table 5 suggests that not all TIMSS items have resulted in a higher mean score at the upper grade level. This pattern exists in all participating nations except for those that gathered data from a single grade (Table 5). The issue of having a higher score at the lower grade level also varies among the nations. In Belgium (Flemish) and South Africa, more than one hundred items have obtained a higher score at the lower grade. On the other hand, in Lithuania, only six items have this problem. The number of the seemingly problematic items for U.S. is 22, less than that of top performing countries, such as Japan (25), Korea (51), and Singapore (24).

Discussion

To disentangle the issue of performance between adjacent grades, features of the TIMSS items should be examined in an international context. Because of the existence of variations in curriculum coverage, development of the TIMSS instrument involves negotiations and compromises among researchers from different participating nations. As a result, not all items fit the assessment needs in each country. Under this general premise, following items can be employed to illustrate some of the technical issues in the mathematics and science benchmarking.

. **Not all questions are developed to fit the curriculum at middle school level**

One of the TIMSS item reads

R12. Subtract: 6000
- 2369

Across all participating nation, no difference was found in the percentage of correct responses (86%) between the adjacent grades in middle school (http://www.timss.org/timss1995i/Items.html). Lower grader in more than one third of the countries (i.e., 16 countries) even had a higher average score on this item. This type of questions may seem too simple for the 7th and 8th graders in most nations. When this item was tried at the elementary school, 71% fourth graders already had the correct answer (http://www.timss.org/timss1995i/Items.html). Therefore, it is fair to conclude that some of the TIMSS items were not designed to reflect the score difference between adjacent grades.

. **Not all items are written with terminologies familiar to international students**

The following is a science item in the TIMSS test:

F01. A small animal called the duckbilled playtypus lives in Australia. Which characteristic of this animal shows that it is a mammal?

While the mammal characteristics have been covered by biology curricula in many nations, it remains unclear whether students in other nations knew the duckbilled playtypus in Australia. The answer could be negative in the United States. The U.S. data indicated that lower graders achieved a higher average score than their peers at the upper grade. In addition, similar problems existed in 10 other countries that had two or more grades participated in the TIMSS test.

. **Not all items provide clear information for precise comprehension**

One of the TIMSS items showed four geometric angles in picture, and questioned: "Which of these angles has a measure closest to 30°?" (Item N15).

Apparently, the word "closest" does not imply "exactly equal to". A careful measure of the four given angles indicated that one of them was exactly equal to 30°. Therefore, this choice may not seem appealing to some students. However, this was the correct answer according to the TIMSS grading code (http://www.timss.org/timss1995i/Items.html). Consequently, one third of the TIMSS participating countries, including the top performing ones like Singapore, Japan, Korea, and Hong Kong, had lower average scores at the upper grade. Similar confusion can be found on another ice-cube question (Item Q18) discussed by Wang (1998).

Despite the existence of various issues in the item construction, it should be acknowledged that TIMSS represents the most extensive comparative study that has ever been undertaken. Thus far, technical concerns have been raised on reporting of the imputed achievement scores under a balanced incomplete block (BIB) design (Wang, 2001). To help reduce the variation of imputed scores within each nation (see Mislevy, 1991; Mislevy, Johnson, & Muraki, 1992), it is important to ensure that the test instrument is appropriate for measuring student achievement in the adjacent grades. In this regard, the preliminary item checking in this investigation may serve as an initial step toward improving the future comparative studies using the TIMSS benchmark.

References

Beaton, A., Martin, M., Mullis, I., Gonzalez, E., Smith, T., & Kelly, D. (1996). Science achievement in the middle school years. Chestnut Hill, MA: Boston College.

Beaton, A., Mullis, I., Martin, M., Gonzalez, E., Kelly, D., & Smith, T. (1996). Mathematics achievement in the middle school years. Chestnut Hill, MA: Boston College.

Lange, J. D. (1997). Looking through the TIMSS mirror from a teaching angle. [On line] Available: http://www.enc.org/topics/timss/additional/documents/0,1341,CDS-000158-cd158,00.shtm (June 11, 2001).

Martin, M. O., & Kelly, D. L. (1996). Technical report volume I: Design and development. Chestnut Hill, MA: TIMSS International Study Center.

Martin, M. O., Mullis, I. V. S., Beaton, A. E., Gonzalez, E. J., Smith, T. A., & Kelly, D. L. (1998). Science achievement in Missouri and Oregon in an international context: 1997 TIMSS benchmarking. Chestnut Hill, MA: TIMSS International Study Center.

Mullis, I. V. S., Martin, M. O., Beaton, A. E., Gonzalez, E. J., Kelly, D. L., & Smith, T. A. (1998). Mathematics achievement in Missouri and Oregon in an international context: 1997 TIMSS benchmarking. Chestnut Hill, MA: TIMSS International Study Center.

Mislevy, R. J. (1991). Randomization-based inference about latent variables from

complex samples. Psychometrika, 56, 177-196.

Mislevy, R., Johnson, E., & Muraki, E. (1992). Scaling procedures in NAEP. Journal of Educational Statistics, 17, 131-154.

Peterson, A. C. (1986, April). Early adolescence: A critical development transition? Paper presented at the 67th annual meeting of the American Educational Research Association, San Francisco, CA.

SPSS (1988). SPSS-X user's guide (3rd ed.). Chicago, IL: Author.

SAS Institute (1990). SAS procedures guide (3rd ed.). Cary, NC: Author.

Schmidt, W. H., McKnight, C. C., Cogan, L. S., Jakwerth, P. M., & Houang, R. T. (1999). Facing the consequences: Using TIMSS for a closer look at U.S. mathematics and science education. Boston, MA: Kluwer.

Walker, E. M., & Madhere, S. (1987). Multiple retentions: Some consequences for the cognitive and affective maturation of minority elementary students. Urban Education, 22 (1), 85-102.

Wang, J. (2001, August-September). TIMSS primary and middle school data: Some technical concerns. Educational Researcher, 17-21.

Wang, J. (1998). A content examination of TIMSS items. Phi Delta Kappan, 80 (1), 36-38.

Table 1

Mean score layout from SAS PROC MEANS

IDCNTRY	IDGRADER	BSMMA01	BSMMA02	BSMMA03
840	low	#	#	#
	up	#	#	#
890	low	#	#	#
	up	#	#	#

Note:

IDCNTRY – country codes
IDGRADER – grade codes
BSMMA01, … – TIMSS item names

Table 2

Partial transpose of mean score layout from SAS PROC MEANS

IDCNTRY	IDGRADER	_Name_	Col1
...			
840	low	BSMMA01	#
	up	BSMMA01	#
890	low	BSMMA01	#
	up	BSMMA01	#
...			
840	low	BSMMA02	#
	up	BSMMA02	#
890	low	BSMMA02	#
	up	BSMMA02	#
...			

Note:

Name a default variable created by SAS to contain item names;
Col1 a default variable created by SAS to contain item scores.

Table 3

SAS statements to compute item score difference between adjacent grades

```
* IDCNTRY – country names;
* IDGRADER – grades;
* TOTWGT – sampling weight;
* BSMMA01 – BSESZ02B (TIMSS item scores);

* (after reading the TIMSS data into SAS);

proc sort;
 by idcntry idgrader;

proc means noprint;
 class idcntry idgrader;
 var BSMMA01--BSESZ02B;
 weight totwgt;
 output out=new(where=(_type_=3)) mean=;

data two;
 set new;
 n=_n_;

proc transpose data=two out=three;
 by n idcntry idgrader;

proc sort;
 by _name_ idcntry idgrader;

data last;
 set three;
 drop n;
 by _name_ idcntry idgrader;
 mean_diff=dif(col1);
 if first.idcntry then mean_diff=.;
 if mean_diff=. then delete;
if mean_diff<0;
proc sort;
by _name_;
proc print;
var IDCNTRY IDGRADER _NAME_ mean_diff;
run;
```

Table 4

A Portion of the SAS output on the item score difference between adjacent grades

The SAS System

Obs	IDCNTRY	IDGRADER	_NAME_	mean_diff
1	56	upper grade	BSEMS01A	-0.001420
2	57	upper grade	BSEMS01A	-0.022135
3	200	upper grade	BSEMS01A	-0.002229
4	344	upper grade	BSEMS01A	-0.015927
5	717	upper grade	BSEMS01B	-0.001311
6	56	upper grade	BSEMS02A	-0.057584
7	608	upper grade	BSEMS02B	-0.007636
8	608	upper grade	BSEMS02C	-0.008507
9	56	upper grade	BSEMT01A	-0.024618
10	608	upper grade	BSEMT01A	-0.001019
...				

Table 5

Number of items resulting in higher average scores at the lower grade level

Country	Number of Items
Australia	12
Austria	34
Belgium (FL)	131
Belgium (FR)	57
Canada	19
Colombia	57
Cyprus	38
Czech	29
Denmark	21
England	25
France	15
Germany	40
Greece	18
Hong Kong	41
Hungary	22
Iceland	43
Iran	60
Ireland	31
Japan	25
Korea	51
Latvia	12
Lithuania	6
Netherlands	58
New Zealand	11
Norway	21
Philippines	81
Portugal	24
Romania	30
Russia	20
Scotland	10
Singapore	24
Slovak	30
Slovenia	28
South Africa	102
Spain	20
Sweden	39
Switzerland	25
Thailand	32
U.S.	22

U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM033762

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: An In-Depth Analysis of Achievement Gaps Between 7th and 8th Grades in the TIMSS Database	
Author(s): Jianjun Wang	
Corporate Source: American Educational Research Association	Publication Date: April 1-5, 2002

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY ______ Sample ______ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY ______ Sample ______ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY ______ Sample ______ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1	Level 2A	Level 2B
[X]	[ ]	[ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate this document as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here,→ please

Signature:	Printed Name/Position/Title: Jianjun Wang / Professor	
Organization/Address: California State University, Bakersfield, CA 93311	Telephone: 661-664-3048	FAX: 661-664-2016
	E-Mail Address: jwang@csub.edu	Date: 3-26-02

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

ERIC CLEARINGHOUSE ON ASSESSMENT AND EVALUATION
UNIVERSITY OF MARYLAND
1129 SHRIVER LAB
COLLEGE PARK, MD 20742-5701
ATTN: ACQUISITIONS

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com